AD-A136 187 ROBUST ESTIMATION BASED ON WALSH AVERAGES FOR THE GENERAL LINEAR MODEL(U) WESTERN MICHIGAN UNIV KALAMAZOO DEPT OF MATHEMATICS G L SIEVERS NOV 83 TR-72 1/1

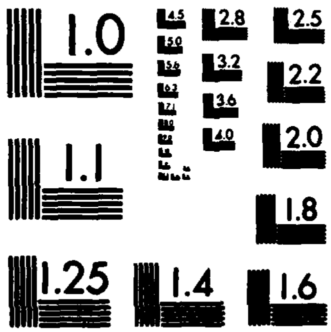
UNCLASSIFIED N00014-78-C-0637 F/G 12/1 NL

END

FILMED

1-84

DTIC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

AD-A136187

ROBUST ESTIMATION BASED ON WALSH AVERAGES
FOR THE GENERAL LINEAR MODEL

by

Gerald L. Sievers

TECHNICAL REPORT NO. 72
November 1983

DEPARTMENT OF MATHEMATICS
WESTERN MICHIGAN UNIVERSITY
Kalamazoo, Michigan
49008

DTIC FILE COPY

DTIC
ELECTE
DEC 21 1983
S D
H

DISTRIBUTION STATEMENT A
Approved for public release;
Distribution Unlimited

83 12 20 120

ROBUST ESTIMATION BASED ON WALSH AVERAGES FOR THE GENERAL LINEAR MODEL

by

Gerald L. Sievers

TECHNICAL REPORT NO. 72
November 1983

This research was supported by the Office of Naval Research under contract N00014-78-C-0637 (NR 042-407)

Reproduction in whole or in part is permitted for any purpose of the United States Government

Approved for public release; distribution unlimited

DEPARTMENT OF MATHEMATICS
WESTERN MICHIGAN UNIVERSITY
KALAMAZOO, MICHIGAN 49008

DTIC COPY INSPECTED 3

Accession For
NTIS GRA&I ☑
DTIC TAB ☐
Unannounced ☐
Justification
By
Distribution/
Availability Codes
Dist | Avail and/or Special
A-1

ABSTRACT

Estimates of the parameters in a linear model are considered based upon the minimization of a dispersion function of the residuals. The dispersion function used depends on Walsh averages of pairs of residuals. Results similar to those arising with signed rank statistics can be obtained as a special case. Trimming and weighting of the Walsh averages can occur with a suitable choice of dispersion function. Asymptotic properties of this type of dispersion function and its derivatives are developed and used to determine the large sample distribution of the estimates. Some discussion appears on the practical application of this methodology.

Key Words and Phrases: M-estimation; Walsh averages; dispersion function; signed rank statistics; robustness

1. INTRODUCTION

This paper is concerned with the development of robust statistical methods based on Walsh averages. The results are broad enough to include many of the familiar results on Walsh averages that arise with signed rank procedures and also to allow for extensions, for instance to trimmed or weighted Walsh averages. The framework of a general linear model is used in the development so that applications can be made to a wide range of statistical problems including one- and two-sample problems, multiple regression problems and analysis of variance and covariance problems. The emphasis will be on the estimation problem although the large sample distributional results can be used to specify tests of hypotheses in a natural way.

The general linear model is given by

$$\underline{Y} = \underline{X}\,\underline{\beta} + \underline{e}, \tag{1.1}$$

where $\underline{Y} = (Y_1, \ldots, Y_n)'$, $\underline{X} = (x_{ij})$ is an $n \times p$ design matrix, $\underline{\beta} = (\beta_1, \ldots, \beta_p)'$ is a $p \times 1$ parameter vector and $e = (e_1, \ldots, e_n)'$ is an $n \times 1$ vector of independent, identically distributed error random variables with density function $f(y)$. It is assumed that $f(y)$ is symmetric about zero. Residuals are denoted by $\underline{Z} = (Z_1, \ldots, Z_n)'$ where

$$\underline{Z} = \underline{Z}(\underline{b}) = \underline{Y} - \underline{X}\,\underline{b}.$$

Consider estimating the parameter $\underline{\beta}$ by minimizing a measure of dispersion of the residuals. In the least-squares approach the sum of squares $\Sigma_i Z_i^2$ is used as the dispersion function. it is well-known that the least-squares estimate is not

robust. It can be inefficient and heavily influenced by outliers in the presence of nonnormal error distributions. The robust M-estimates developed by Huber (1964, 1972, 1973) arise by minimizing a dispersion function $\Sigma_i \rho(Z_i)$ for a suitably chosen convex function ρ. To attain a degree of robustness the function should increase at a lesser rate than the quadratic function in its tails. The ℓ_1 or least absolute deviation method minimizes $\Sigma_i |Z_i|$. The dispersion function $\Sigma_i\ a(R_i^+)|Z_i|$, where $a(\cdot)$ is a score function and R_i^+ is the rank of Z_i in absolute value, generates an estimate of $\underline{\beta}$ based on signed-rank statistics, see Adichie (1967, 1978).

The basic dispersion function to be considered here measures variability in the Walsh averages of the residuals with

$$D = D(\underline{b}) = \Sigma_{i<j}\ w_{ij}\rho(Z_i + Z_j), \tag{1.2}$$

where ρ is a convex function. For convenience, the "2" in the denominator of Walsh averages has been absorbed in the ρ function. The constants $w_{ij} \geq 0$ are weights reflecting the importance of individual Walsh averages. The weights can depend on the design matrix. Zero-one weights can be used to omit some Walsh averages from consideration.

For the present, three potential ρ functions will be mentioned. The first is simply

$$\rho_1(t) = |t|. \tag{1.3}$$

If this ρ function is used in (1.2) with weights $w_{ij} \equiv 1$, the dispersion function is very similar to that of the signed rank approach with Wilcoxon scores. For example, in the

one-sample problem with $Y_i = \theta + e_i$ the dispersion function using $\rho_1(t)$ is minimum at the median of the Walsh averages $(Y_i + Y_j)/2$, for $i < j$, which is essentially the signed rank estimate of θ .

Another ρ function is

$$\rho_2(t) = \max\{|t| - c,\ 0\} \tag{1.4}$$

for some $c > 0$. This function is zero on the interval $[-c, c]$ and in effect trims "middle" Walsh averages that are sufficiently near zero. However, $\rho_2(t)$ can also be viewed as a simple modification of $\rho_1(t)$ which flattens its abrupt behavior at $t = 0$. A consequence of this modification may be that the standard error of $\hat{\underline{\beta}}$ becomes more stable, but this conjecture needs further examination.

Huber's ρ function can also be used in the dispersion (1.2). It is quadratic in the middle with linear tails and is given by

$$\rho_3(t) = \begin{cases} t^2/2 & \text{if } |t| \le k \\ k|t| - k^2/2 & \text{if } |t| > k, \end{cases} \tag{1.5}$$

for some $k > 0$.

The above ρ functions suggest what might be accomplished by the use of the dispersion function (1.2). With $\rho_1(t)$ and no weights the estimate should be similar to that arising with the signed rank dispersion function. The use of weights allows broader possibilities and the modification to $\rho_2(t)$ may prove useful. On the other hand, the use of $\rho_3(t)$ suggests this to be an extension of the M-estimate approach (Huberizing Walsh

averages). Huber (1964) had mentioned this type of idea at the end of his first paper.

2. THE MAIN RESULTS

In this section the basic notation is introduced and the assumptions are listed. The basic focus will be on the derivative of the dispersion function (1.2). Theorem 2.1 shows that this derivative has a multivariate normal limiting distribution. This is extended to the case of contiguous distributions in Theorem 2.4. These results are useful in developing test statistics for testing hypotheses about $\underline{\beta}$. An asymptotic linearity result is given in Theorem 2.3 and this is used to drive the limiting normality of the estimate $\hat{\underline{\beta}}$ in Theorem 2.5.

Some assumptions will be listed concerning the design matrix $\underline{X}$ and the weights used in the dispersion function. Extend the definition of the weights to the case of $i > j$ by defining $w_{ji} = w_{ij}$. Also let $w_{ii} = \Sigma_{j \neq i}\, w_{ij}$ and define an $n \times n$ weight matrix $\underline{W} = (w_{ij})$. Then $\underline{W}$ is a symmetric matrix with weights w_{ij} in the off-diagonal locations and its diagonal elements are the sums of the off-diagonal elements in the corresponding row. Further, define $a_{ij}(k) = w_{ij}(x_{ik} + x_{jk})$ and let $A_i(k) = \Sigma_{j \neq i}\, a_{ij}(k)$. A calculation shows that the $n \times p$ matrix having $A_i(k)$ for its i,k^{th} element is given by $\underline{A} = \underline{W}\,\underline{X}$.

<u>ASSUMPTION (A_1):</u>

$$(1/n)\underline{X}'\underline{X} \quad \to \quad \underline{\Sigma}$$

as $n \to \infty$, where $\underline{\Sigma}$ is a $p \times p$ positive definite matrix.

ASSUMPTION (A_2): For each $k = 1, \ldots, p$

$$\max_{1 \leq i \leq n} |x_{ik}| / \sqrt{n} \to 0 \quad \text{as} \quad n \to \infty \quad .$$

ASSUMPTION (A_3): For each $k = 1, \ldots, p$

$$\frac{\max_{1 \leq i \leq n} A_i^2(k)}{\sum_{i=1}^{n} A_i^2(k)} \to 0 \quad \text{as} \quad n \to \infty \; .$$

ASSUMPTION (A_4): For each $k = 1, \ldots, p$

$$\frac{\sum_{i<j} a_{ij}^2(k)}{\sum_{i=1}^{n} A_i^2(k)} \to 0 \quad \text{as} \quad n \to \infty \; .$$

ASSUMPTION (A_5):

$$n^{-3} \underline{X}' \, \underline{W} \, \underline{W} \, \underline{X} \quad \to \quad \underline{V}$$

as $n \to \infty$, where $\underline{V}$ is a $p \times p$ positive definite matrix.

ASSUMPTION (A_6):

$$n^{-2} \underline{X}' \, \underline{W} \, \underline{X} \quad \to \quad \underline{C}$$

as $n \to \infty$, where $\underline{C}$ is a $p \times p$ nonsingular matrix.

The following assumptions concern the ρ function in the dispersion (1.2). They are sometimes motivated by the approach used in the proofs and alternate assumptions could be specified

with different proof techniques. The basic requirement is that ρ and its derivative be sufficiently smooth(piecewise). In many cases of practical interest ρ has a bounded, piecewise continuous derivative and then the assumptions could be considerably simplified.

ASSUMPTION (B_1): $\rho(t)$ is a convex function, symmetric about zero, with a derivative $\psi(t) = \rho'(t)$ except at possibly a finite number of points. This implies that $\psi(t)$ is nondecreasing and $\psi(-t) = -\psi(t)$.

ASSUMPTION (B_2): $h(t) = E_0(\psi(Y_1 + t))$ exists and is finite for all t, where the expectation is under the assumption that $\underline{\beta} = \underline{0}$ in model (1.1).

ASSUMPTION (B_3): The following expectations are positive and finite:

$$\tau^2 = E_0(h^2(Y_1)) = E_0(\psi(Y_1 + Y_2)\psi(Y_1 + Y_3)),$$

$$\tau_1^2 = E_0(\psi^2(Y_1 + Y_2)) \quad \text{and}$$

$$E((h'(Y_1))^2).$$

ASSUMPTION (B_4): The first and second derivatives $h'(t)$ and $h''(t)$ exist except possibly at a finite number of points and $|h''(t)| \leq M$ for some constant M.

ASSUMPTION (B_5): $H(t) = E_0(h(Y_1 - t)) = E_0(\psi(Y_1 + Y_2 - t))$ and its derivative exist in a neighborhood of $t = 0$. Moreover, $H'(t)$ is continuous at $t = 0$ and $H'(0) = -E_0(h')$.

ASSUMPTION (B_6): For some constant M_1, $E_0(\psi^2(Y_1 + Y_2 - t)) \leq M_1$ in some neighborhood of $t = 0$.

The behavior of the dispersion function (2.1) can be studied through the vector of its derivatives. The negatives of these derivatives will be denoted by $\underline{T}(\underline{b}) = (T_1(\underline{b}), \ldots, T_p(\underline{b}))'$ where

$$T_k(\underline{b}) = -\partial D/\partial b_k = \Sigma_{i<j}\, a_{ij}(k)\psi(Z_i + Z_j) \qquad (2..$$

for $k = 1, \ldots, p$, where $\psi = \rho'$ and $a_{ij}(k) = w_{ij}(x_{ik} + x_{jk})$.

The asymptotic distribution of $\underline{T}(\underline{b})$ will be treated by the projection method. It will be sufficient to assume $\underline{\beta} = \underline{0}$ in model (1.1), in which case the Y_i are iid with symmetric density f. The k^{th} coordinate of $\underline{T}(\underline{0})$ has projection

$$\begin{aligned} T_k^*(0) &= \textstyle\sum_{\ell=1}^n E_0(T_k(\underline{0}) \mid Y_\ell = y_\ell) \\ &= \textstyle\sum_{\ell=1}^n \sum_{i<j} a_{ij}(k) E_0(\psi(Y_i + Y_j) \mid Y_\ell = y_\ell) \qquad (2.2) \\ &= \textstyle\sum_{\ell=1}^n A_\ell(k) h(Y_\ell), \end{aligned}$$

where $h(t)$ is defined in assumption (B_2). Note that $E_0(\psi(Y_i + Y_j)) = 0$ was used since ψ is an odd function under assumption (B_1). Thus the projection of $\underline{T}(\underline{0})$ is $\underline{T}^*(\underline{0}) = (T_1^*(\underline{0}), \ldots, T_p^*(\underline{0}))'$. In matrix form,

$$\underline{T}^*(\underline{0}) = \underline{A}'\,\underline{H} = \underline{X}'\,\underline{W}\,\underline{H}$$

where $\underline{H} = (h(Y_1), \ldots, h(Y_n))'$.

<u>THEOREM 2.1.</u> Let assumptions (A_3) - (A_5) and (B_1) -(B_3) hold. Then for any fixed vector $\underline{\theta} = (\theta_1, \ldots, \theta_p)'$,

$$\text{(i)} \quad n^{-3/2}(\underline{\theta}'\underline{T}(\underline{0}) - \underline{\theta}'\underline{T}^*(\underline{0}))|_{\underline{0}} \xrightarrow{P} 0 \quad \text{and}$$

$$\text{(ii)} \quad n^{-3/2}\underline{T}(\underline{0})|_{\underline{0}} \xrightarrow{L} N(\underline{0}, \tau^2\underline{V}) \quad \text{as } n \to \infty$$

Proof: First let

$$\underline{u} = \underline{W}\,\underline{X}\,\underline{\theta} = (u_1, \ldots, u_n)'. \tag{2.3}$$

Then $n^{-3/2}\underline{\theta}'\underline{T}^*(\underline{0}) = n^{-3/2}\underline{u}'\,\underline{H}$ is a sum of independent random variables with mean 0 and variance $n^{-3}\,\tau^2\,\Sigma_i\,u_i^2 =$ $n^{-3}\,\tau^2\,\underline{\theta}'\,\underline{X}'\,\underline{W}\,\underline{W}\,\underline{X}\,\underline{\theta} \to \tau^2\,\underline{\theta}'\,\underline{V}\,\underline{\theta}$ by assumption (A_5). It will have a limiting normal distribution if $\max_{1\leq i\leq n} u_i^2/\sum_{i=1}^n u_i^2 \to$ 0 as $n \to \infty$. But this follows from assumptions (A_3) and (A_5). Thus $n^{-3/2}\underline{T}^*(\underline{0})|_{\underline{0}} \xrightarrow{L} N(0, \tau^2\underline{V})$ and part (ii) will follow from part (i).

For part (i) examine the expected square

$$\begin{aligned} n^{-3}E_0(\underline{\theta}'\underline{T}(\underline{0}) &- \underline{\theta}'\underline{T}^*(\underline{0}))^2 \\ &= n^{-3}(E_0(\underline{\theta}'\underline{T}(\underline{0}))^2 - E_0(\underline{\theta}'\underline{T}^*(\underline{0}))^2) \\ &= (\tau_1^2 - 2\tau^2)(n^{-3}\underline{u}'\underline{u})(\textstyle\sum_{i<j} u_{ij}^2/\sum_i u_i^2), \end{aligned}$$

where $u_{ij} = \sum_k \theta_k\, a_{ij}(k)$. The middle factor converges to a constant as in the previous paragraph and the last factor tends to zero by assumptions (A_4) and (A_5). Thus part (i) follows.

THEOREM 2.2. Let assumptions (A_1), $(A_3)-(A_5)$ and $(B_1)-(B_4)$ hold. Let $\underline{\Delta} = (\Delta_1, \ldots, \Delta_p)'$ and

$$\underline{R}(\underline{\Delta}) = n^{-3/2}(\underline{T}(\underline{\Delta}/\sqrt{n}) - \underline{T}(\underline{0}) + E_0(h')\underline{X}'\underline{W}\,\underline{X}\,\underline{\Delta}/\sqrt{n}).$$

Then $\underline{R}(\underline{\Delta})|_{\underline{0}} \xrightarrow{P} \underline{0}$ as $n \to \infty$.

Proof: First extend the definition of $\underline{T}^*(\underline{0})$ in (2.2) by defining $\underline{T}^*(\underline{b})$ to have k^{th} element

$$T_k^*(\underline{b}) = \sum_{i=1}^{n} A_i(k)h(Z_i)$$

where $\underline{Z} = \underline{Y} - \underline{X}\,\underline{b}$. Note that $\underline{T}(\underline{b})$ has the translation property $\underline{T}(\underline{b}_1)|_{\underline{b}_2} \stackrel{L}{=} \underline{T}(\underline{b}_1 - \underline{b}_2)|_{\underline{0}} \stackrel{L}{=} \underline{T}(\underline{0})|_{\underline{b}_2 - \underline{b}_1}$ and so also does $\underline{T}^*(\underline{b})$. Then Theorem 2.1 (i) and a contiguity argument shows that

$$n^{-3/2}(\underline{\theta}'\underline{T}(\underline{0}) - \underline{\theta}'\underline{T}^*(\underline{0}))|_{-\underline{\Delta}/\sqrt{n}} \xrightarrow{P} 0.$$

Using the translation property it follows that

$$n^{-3/2}(\underline{\theta}'\underline{T}(\underline{\Delta}/\sqrt{n}) - \underline{\theta}'\underline{T}^*(\underline{\Delta}/\sqrt{n}))|_{\underline{0}} \xrightarrow{P} 0.$$

Thus it is sufficient to replace $\underline{T}$ by $\underline{T}^*$ in verifying that $\underline{\theta}'\underline{R}(\underline{\Delta})|_{\underline{0}} \xrightarrow{P} 0$.

Define an $n \times 1$ vector of constants $\underline{t} = \underline{X}\,\underline{\Delta}/\sqrt{n}$. Then with $\underline{u}$ as in (2.3) use a Taylor's approximation to write

$$\underline{\theta}'\underline{T}^*(\underline{\Delta}/\sqrt{n}) = \sum_{i=1}^{n} u_i h(Y_i - t_i)$$
$$= \sum_i u_i h(Y_i) - \sum_i u_i t_i h'(Y_i)$$
$$+ \sum_i u_i t_i^2 h''(\xi_i)/2$$

where $|\xi_i - Y_i| \leq t_i$, $i = 1, \ldots, n$. Then

$$\begin{aligned}\underline{\theta}'\underline{R}^*(\underline{\Delta}) &= n^{-3/2}(\underline{\theta}'\underline{T}^*(\underline{\Delta}/\sqrt{n}) - \underline{\theta}'\underline{T}^*(\underline{0}) + E_0(h')\underline{\theta}'\underline{X}'\underline{W}\ \underline{X}\ \underline{\Delta}/\sqrt{n}) \\ &= n^{-3/2}(\Sigma_i\ u_i h(Y_i + t_i) - \Sigma_i\ u_i h(Y_i) + E_0(h')\Sigma_i\ u_i t_i) \\ &= -n^{-3/2}\ \Sigma_i\ u_i t_i (h'(Y_i) - E_0(h')) + n^{-3/2}\ \Sigma_i\ u_i t_i^2 h''(\xi_i)/2 \\ &= S_1 + S_2 \quad \text{say.}\end{aligned}$$

Now S_1 is a sum of independent random variables with mean zero and variance

$$\begin{aligned}&n^{-3}\ \Sigma_i\ u_i^2 t_i^2\ \mathrm{Var}(h'(Y_1)) \\ &\qquad \leq (n^{-3}\ \underline{u}'\underline{u})\ (\max_{1\leq i\leq n}\ u_i^2/\Sigma_i\ u_i^2)(\Sigma_i t_i^2)\ \mathrm{Var}(h'(Y_1)).\end{aligned}$$

The first factor here converges to a constant and the second factor converges to zero as in the proof of Theorem 2.1. Also $\Sigma_i\ t_i^2 = \underline{\Delta}'\underline{X}'\underline{X}\ \underline{\Delta}/n \to \underline{\Delta}'\underline{\Sigma}\ \underline{\Delta}$ by assumption (A_1). With its variance tending to zero, $S_1 \xrightarrow{P} 0$. Using the bound on h'' in assumption (B_4), the term S_2 is bounded by a constant which tends to zero,

$$|S_2| \leq n^{-3/2} \max_{1\leq i\leq n} |u_i|\ M\ \Sigma_i\ t_i^2/2.$$

Thus $S_2 \xrightarrow{P} 0$ and the proof is completed.

The previous theorem shows that $\underline{T}(\underline{b})$ can be approximated by a linear function of $\underline{b}$ for $\underline{b}$ near zero. However, the result is not strong enough for the application needed here. The following theorem shows that this result holds uniformly. A proof will not be given as it is quite lengthy and the details follow closely the compactification argument used in the proof of Theorem 5.1 of Sievers (1983).

<u>THEOREM 2.3</u>. Let assumptions (A_1) - (A_5) and (B_1) -(B_6) hold. Let $\mathcal{D} = \{(\Delta_1, \ldots, \Delta_p): |\Delta_k| \leq c, 1 \leq k \leq p\}$, where $c > 0$, and let $||\cdot||$ denote Euclidean distance. Then

$$\sup_{\underline{\Delta} \in \mathcal{D}} ||\underline{R}(\underline{\Delta})|| \xrightarrow{P} 0 \quad \text{as} \quad n \to \infty \ .$$

The asymptotic distribution of $\underline{T}(\underline{0})$ given in Theorem 2.1 can be extended to the case of contiguous distributions. The result follows readily from Theorem 2.2 and is summarized in the following theorem.

<u>THEOREM 2.4</u>. Let assumptions (A_1), (A_3) - (A_6) and (B_1) - (B_4) hold. Then as $n \to \infty$,

$$n^{-3/2}\underline{T}(\underline{0})|_{\underline{\Delta}/\sqrt{n}} \xrightarrow{\mathcal{L}} N(E_0(h')\underline{C}\underline{\Delta}, \tau^2\underline{V}).$$

Finally, the limiting distribution of the estimate $\hat{\underline{\beta}}$ can be given. With the asymptotic linearity result of Theorem 2.3, the argument of Jaeckel (1972) and Sievers (1983) can be used.

First note a translation property of the estimate, $\sqrt{n}(\hat{\underline{\beta}} - \underline{\beta})|_{\underline{\beta}} \xrightarrow{L} \hat{\underline{\Delta}}|_{\underline{0}}$, where $\hat{\underline{\Delta}}$ minimizes $D^*(\underline{\Delta}) = D(\underline{\Delta}/\sqrt{n})/n$. The asymptotic linearity implies $\sup_{\underline{\Delta} \in \mathcal{D}} |D^*(\underline{\Delta}) - Q(\underline{\Delta})||_{\underline{0}} \xrightarrow{P} 0$, where Q is the quadratic function

$$Q(\underline{\Delta}) = E_0(h')\underline{\Delta}'\underline{C}\,\underline{\Delta}/2 - n^{-3/2}\underline{\Delta}'\underline{T}(\underline{0}) + D^*(\underline{0}).$$

Form this it follows that $\hat{\underline{\Delta}}$ is asymptotically equivalent to the point minimizing $Q(\underline{\Delta})$. The following theorem summarizes.

<u>THEOREM 2.5</u>. Let assumptions $(A_1) - (A_6)$ and $(B_1) - (B_6)$ hold. Then as $n \to \infty$

$$\sqrt{n}(\hat{\underline{\beta}} - \underline{\beta})|_{\underline{\beta}} \xrightarrow{L} N(\underline{0}, (\tau/E_0(h'))^2\underline{C}^{-1}\underline{V}\,\underline{C}^{-1}).$$

3. GENERAL COMMENTS

The regular M-estimate of $\underline{\beta}$ minimizing $\sum\rho(Z_i)$ has an influence function proportional to $\psi(y)$ and its asymptotic variance-covariance matrix is $E(\psi^2)/(E(\psi'))^2\ \underline{\Sigma}^{-1}$. The estimate of $\underline{\beta}$ minimizing the dispersion (1.2) has an influence function $h(y)$, which is a smoothed version of $\psi(y)$, and its variance-covariance matrix, given in Theorem 2.5, may have a factor larger or smaller than that of the regular M-estimate. Some examples of these quantities appear in the next section.

There is special interest in conditions under which the matrix $\underline{C}^{-1}\underline{V}\ \underline{C}^{-1}$, appearing in Theorem 2.5, equals $\underline{\Sigma}^{-1}$. If this is the case, the variances of $\hat{\underline{\beta}}$ can be compared to the variances of regular M-estimates and least-squares estimates simply by the constant multiples of this matrix. An answer to this question can be given for the unweighted case, $w_{ij} \equiv 1$. In this case $\underline{W} = (n-2)\underline{I} + \underline{J}$, where $\underline{I}$ is an identity matrix and $\underline{J}$ a matrix of "ones". Then $\underline{C} = \underline{\Sigma} + \underline{\mu}\ \underline{\mu}'$ and $\underline{V} = \underline{\Sigma} + 3\underline{\mu}\ \underline{\mu}'$, where $\underline{\mu}$ is the limit of the column means of $\underline{X}$. Then a sufficient condition for $\underline{C}^{-1}\ \underline{V}\ \underline{C}^{-1} = \underline{\Sigma}^{-1}$, equivalently $\underline{V} = \underline{C}\ \underline{\Sigma}^{-1}\ \underline{C}$, is given by

$$\underline{\mu}'\underline{\Sigma}^{-1}\underline{\mu} = 1, \tag{3.1}$$

as can be seen by direct multiplication. This condition is easier to verify in particular cases than the basic equation itself.

To estimate the standard errors of the estimates in $\hat{\underline{\beta}}$ an estimate is needed for the scale factor $\tau^2/(E_0(h'))^2$. Recalling

that $\tau^2 = E_0(\psi(Y_1 + Y_2)\psi(Y_1 + Y_3))$, a U-statistic is suggested for the numerator. A symmetric kernel is given by $\phi(Y_1, Y_2, Y_3) = (\psi(Y_1 + Y_2)\psi(Y_1 + Y_3) + \psi(Y_1 + Y_2)\psi(Y_2 + Y_3) + \psi(Y_1 + Y_3)\psi(Y_2 + Y_3))/3$. Then using residuals $\hat{\underline{Z}} = \underline{Y} - \underline{X}\,\hat{\underline{\beta}}$, a consistent estimate of τ^2 is

$$\hat{\tau}^2 = \textstyle\sum_{i<j<k}\phi(\hat{Z}_i, \hat{Z}_j, \hat{Z}_k)/\binom{n}{3} .$$

To estimate the denominator of the scale factor consider the case where $h'(t) = \int\psi'(y + t)f(y)dy$. Then $E_0(h') = E_0(\psi'(Y_1 + Y_2))$ and a consistent, U-statistic estimate is given by $\sum_{i<j} \psi'(Z_i + Z_j)/\binom{n}{2}$.

The computational aspects discussed in Huber (1972, 1973) for regular M-estimates could be modified for use in computing $\hat{\underline{\beta}}$ for ρ functions satisfying his conditions. In particular, a scale measure should be used with some ρ functions, such as (1.5). The process will be slower since the dispersion (1.2) involves $\binom{n}{2}$ rather than n terms.

There is another type of dispersion function that can be used for the analysis of a linear model. Consider

$$D_1 = D_1(\underline{b}) = \textstyle\sum_{i<j} w_{ij}\ (Z_j - Z_i).$$

Dispersion is measured by differences of residuals. This is a generalization of the dispersion function considered in Sievers (1983) where $\rho(t) = |t|$ was used. With no weights this Gini's mean difference was shown by Hettmansperger and McKean (1978) to generate the rank estimate of $\underline{\beta}$ based on Wilcoxon scores. The

projection and asymptotic linearity approach of this paper can be used with only minor changes to obtain the theoretical properties for the estimate minimizing D_1. The results are basically the same as Theorems 2.1 - 2.5 with some changes in the details.

Tests of hypotheses can be developed based on $\underline{T}(\underline{0})$, $\hat{\underline{\beta}}$ or the dispersion function, see Hettmansperger and McKean (1977) and Schrader and Hettmansperger (1980).

4. EXAMPLES

The introduction discussed three possible ρ functions for use in the dispersion (1.2). Further details on these functions will be given in this section, in particular, on the influence function $h(y)$ and quantities appearing in the asymptotic variance. Some comments are made on the one- and two-sample problems and on the simple linear regression model.

The function $\rho_1(t) = |t|$ has derivative

$$\psi_1(t) = \begin{cases} +1 & \text{if } t > 0 \\ -1 & \text{if } t < 0. \end{cases}$$

Then the influence function is $h(t) = 2F(t) - 1$ and $\tau^2 = 1/3$. Also $h'(t) = 2f(t)$ and $E_0(h') = 2\int f^2$. The asymptotic variance factor is $1/12(\int f^2)^2$, the familiar result for signed rank estimates.

The function $\rho_2(t)$ of (1.4) has derivative

$$\psi_2(t) = \begin{cases} -1 & \text{if } t < -c \\ 0 & \text{if } |t| \leq c \\ +1 & \text{if } t > c. \end{cases}$$

Then the influence function is $h(t) = F(c + t) - F(c - t)$ and τ^2 is the expected square of this function. Also $h'(t) = f(c + t) + f(c - t)$. The expected value of h' can be expressed as $E_0(h') = 2g(c)$, where $g(y)$ is the density function of $Y_1 + Y_2$.

The Huber function $\rho_3(t)$ of (1.5) has derivative

$$\psi_3(t) = \begin{cases} -k & \text{if } t < -k \\ t & \text{if } |t| \leq k \\ +k & \text{if } t > k. \end{cases}$$

The influence function is $h(t) = \int_{t-k}^{t+k} F(u)du - k$ and it can be viewed as a smoothing of ψ_3. It readily can be seen that $h'(t) = F(t + k) - F(t - k)$ and $E_0(h') = P_0(|Y_1 + Y_2| \leq k)$.

For the one-sample problem the Y_i are assumed to be symmetric about a point θ and the dispersion function is $D(\theta) = \sum_{i<j} \rho(Y_i + Y_j - 2\theta)$. There appears to be no use for weights here. The assumptions simplify considerably. If ρ_1 is used, the estimate is the median of $(Y_i + Y_j)/2$ for $i < j$. The effect of the Walsh averages on $\hat{\theta}$ can be trimmed or smoothed with other choices of the ρ function.

For the two-sample problem suppose there are samples of sizes n_1 and n_2 from two groups G_1 and G_2 with locations β_1 and β_2. Write the design matrix as

$$\underline{X} = \begin{pmatrix} \underline{1} & \underline{0} \\ \underline{0} & \underline{1} \end{pmatrix} ,$$

where $\underline{0}$ and $\underline{1}$ are vectors of zeros and ones, respectively. Suppose $w_{ij} = w_{11}$ if i,j are both in G_1, $w_{ij} = w_{22}$ if i,j are both in G_2 and $w_{ij} = w_{12}$ if i,j are in different groups. Assumptions (A_1) - (A_6) will hold in this case. The dispersion function becomes

$$D = w_{11}\ \Sigma_{G_1,G_1} \rho(Y_i + Y_j - 2\beta_1) + w_{22}\ \Sigma_{G_2,G_2} \rho(Y_i + Y_j - 2\beta_2)$$

$$+ w_{12}\ \Sigma_{G_1,G_2}\ \rho(Y_i + Y_j - \beta_1 - \beta_2).$$

It appears that $\hat{\beta}_1$ depends on the data from both groups if $w_{12} \neq 0$; similarly for $\hat{\beta}_2$. This differs from the regular M-estimate method where $\hat{\beta}_k$ depends only on the data from group G_k, $k = 1, 2$. It can be verified by direct computation that $\underline{C}^{-1}\underline{V}\ \underline{C}^{-1} = \underline{\Sigma}^{-1}$. This is so regardless of the choice of w_{11}, w_{22} and w_{12} and as a result these weights have no effect on the asymptotic variances.

In the simple linear regression model $Y_i = \beta_1 + \beta_2 x_i + e_i$, $1 \leq i \leq n$. The dispersion function is

$$D = \Sigma_{i<j}\ w_{ij}\ \rho(Y_i + Y_j - 2\beta_1 - (x_i + x_j)\beta_2).$$

In the case of no weights, $w_{ij} \equiv 1$, expressions for $\underline{C}$, $\underline{V}$ and $\underline{\Sigma}$ are readily computed and (3.1) implies $\underline{C}^{-1}\underline{V}\ \underline{C}^{-1} = \underline{\Sigma}^{-1}$, the familiar matrix for this problem. It is not clear if this can hold for other choices of weights.

REFERENCES

Adichie, J.N. (1967), "Estimates of Regression Parameters Based on Rank Tests, "Annals of Mathematical Statistics, 38, 894-904.

Adichie, J.N. (1978), "Rank Tests of Sub-Hypotheses in the General Linear Regression," Annals of Statistics, 6, 1012-1026.

Hettmansperger, T.P. and McKean, J.W. (1977), "A Robust Alternative Based on Ranks to Least Squares in Analyzing Linear Models," Technometrics, 19, 275-284.

Hettmansperger, T.P. and McKean, J.W. (1978), "Statistical Inference Based on Ranks," Psychometrika, 43, 69-79.

Huber, P.J. (1964), "Robust Estimation of a Location Parameter," Annals of Mathematical Statistics, 35, 73-101.

Huber, P.J. (1972), "Robust Statistics," Annals of Mathematical Statistics, 43, 1041-1067.

Huber, P.J. (1973), "Robust Regression: Asymptotics, Conjectures and Monte Carlo," The Annals of Statistics, 1, 799-821.

Jaeckel, L.A. (1972), "Estimating Regression Coefficients By Minimizing the Dispersion of the Residuals," Annals of Mathematical Statistics, 43, 1449-1458.

Schrader, R.M. and McKean, J.W. (1980), "Robust Analysis of Variance Based Upon a Likelihood Criterion," Biometrika, 67, 93-101.

Sievers, G.L. (1983), "A Weighted Dispersion Function For Estimation In Linear Models," Communications in Statistics -Theory and Methods - A, 12, 1161-1179.

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER Technical Report #72	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Robust Estimation Based on Walsh Averages For the General Linear Model		5. TYPE OF REPORT & PERIOD COVERED Technical Report 1983
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Gerald L. Sievers		8. CONTRACT OR GRANT NUMBER(s) N 00014-78-C-0637
9. PERFORMING ORGANIZATION NAME AND ADDRESS Western Michigan University Kalamazoo, Michigan 49008		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NR 042-407
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research Statistics and Probability Program		12. REPORT DATE November 1983
		13. NUMBER OF PAGES 22
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved For Public Release: Distribution Unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

M-estimation; Walsh averages; dispersion function; signed rank statistics; robustness

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Please see next page

20. ABSTRACT

Estimates of the parameters in a linear model are considered based upon the minimization of a dispersion function of the residuals. The dispersion function used depends on Walsh averages of pairs of residuals. Results similar to those arising with signed rank statistics can be obtained as a special case. Trimming and weighting of the Walsh averages can occur with a suitable choice of dispersion function. Asymptotic properties of this type of dispersion function and its derivatives are developed and used to determine the large sample distribution of the estimates. Some discussion appears on the practical application of this methodology.

FILM

1-8